Privacy issues of ISPs in the modern web

Privacy Issues of ISPs in the Modern Web

Ali Safari Khatouni, Martino Trevisan, Leonardo Regano, Alessio Viticchié
Politecnico di Torino, Italy
firstname.lastname@polito.it

***Abstract*—In recent years, privacy issues in the networking field are getting more important. In particular, there is a lively debate about how Internet Service Providers (ISPs) should collect and treat data coming from passive network measurements. This kind of information, such as flow records or HTTP logs, carries considerable knowledge from several points of view: traffic engineering, academic research, and web marketing can take advantage from passive network measurements on ISP customers. Nevertheless, in many cases collected measurements contain personal and confidential information about customers exposed to monitoring, thus raising several ethical issues. Modern web is very different from the one we experienced few years ago: web services converged to few protocols (i.e., HTTP and HTTPS) and a large share of traffic is encrypted.**

The aim of this work is to provide an insight about which information is still visible to ISPs, with particular attention to novel and emerging protocols, and to what extent it carries personal information. We illustrate that sensible information, such as website history, is still exposed to passive monitoring. We illustrate privacy and ethical issues deriving by the current situation and provide general guidelines and best practices to cope with the collection of network traffic measurements.

Keywords—Passive Monitoring, Privacy, ISP

I. INTRODUCTION

Passive measurements are the most practical means to measure the behavior of network users. They provide immediate and detailed insights about the usage of the network at the physical layer; furthermore, when suitable processing is performed on collected data, it is possible to extract higher level metrics to measure, e.g., users' perceived Quality of Experience (QoE) [1], video streaming quality, etc. The collected data contains knowledge about the users and services that they are using. Thus, it possibly exposes private information or user's credentials. Privacy and security issues are strongly related to the adopted protocols, if they are encrypted or they transmit users credentials as clear text, etc. Several ethical issues arise when ISP and network administrators cope with passive measurements, and it is not often clear to what extent user's privacy is broken e.g., i) which kind of data is visible by network probes, ii) what can be extracted from the user's flow records, iii) how should the extracted data be stored, iv) who has the right to access the collected data, etc. Many works investigate about the ethical arguments in the modern web; in particular Zevenbergen et al. [2] and Vassio et al. [3] discuss the ethical issues related to traffic measurement, and provide guidelines on how to deal with personal data.

To illustrate the considered scenario, Fig. 1 shows a typical use case of network passive monitoring. A probe seats at ISP level, e.g., a Point of Presence (PoP) where households' traffic is aggregated. All users' connections behind the probe can be possibly captured and analyzed by the probe. Moreover, nowadays probes are able to filter connections or packets with protocol-based and content-based filtering rules. The fundamental questions are: i) which information is visible at the ISP level, ii) how much information can be captured or stored at the probes. The storage security and accessibility of the collected data are out of scope of this work since several generic and well-known techniques exist. However, we deal with possible ethical and privacy issues about extracted data from users.

A passive probe captures and analyzes traffic, grabbing (or sniffing) packets transmitted on network cable(s). Each packet is composed by several protocol headers and a payload. The headers contain useful information such IP addresses and ports of the two communication endpoints. Typically, network packets are aggregated into *flows* (TCP or UDP connections) by grouping them by IP address and port number. All packets belonging to the same flow are analyzed together to compute overall statistics; the set of techniques to study network flows is called *flow monitoring* [4]. Each flow record includes fine grained information about the flow start time, duration, and byte transfer by endpoints. ISPs have personal information about all individual customers from the moment they apply for the service and, furthermore, they are responsible for managing and assigning IP addresses. Thus, it is trivial for IPSs to associate traffic to the single customer by simply looking up the client IP address of flows. Beside the customer identifier, passive measurements give information about the contacted services, at which time and with which order they are contacted. The server IP address provides information about the server location and, possibly, about the services provided by it. That is, the advent of Content Delivery Networks (CDN) and Cloud Providers made the Internet tangle in a scenario where the same server can offer several services and web pages based on user geographic location and more sophisticated traffic engineering rules.

CDNs and Cloud Providers are an example of the significant change of the Internet in the last decade. The aim of this work is to summarize the evolution of the Internet in terms of traffic volumes and protocols and to debate on the potential ethical and moral issues regarding the use of collected network data at ISP level.

The main contributions of the paper are:

- Quantification of the changes in protocol usage and traffic volumes on Internet in the recent years (Section III).

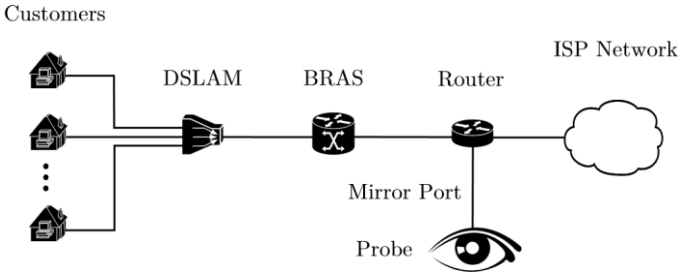

Fig. 1: Typical network probe deployment at ISP level

- Analysis of the personal information carried by network packet, separately by ISO/OSI layer, and focusing on emerging protocols (Section IV).
- Description of the ethical issues that arise from network monitoring (Section V).
- Proposal of alternative scenarios that offer a reasonable trade-off between users' privacy and ISPs' greed of pervasive network measurements (Section VI).

The paper is organized as follows. Section II shortly overviews the involved stakeholders in the current Internet ecosystem. Section III describes the recent evolution of Internet traffic. Section IV discusses how personal information can be monitored at the ISP level, and section V presents the privacy and ethical issues that arise with passive network measurement. Section VI defines our alternative proposal. Finally, Section VII summarizes our findings.

II. Involved Stakeholders

The Internet ecosystem is a very tangled and complex universe. Since many actors are involved with different interests, in this section we provide an in-depth analysis of the *stakeholders* acting in this system.

The main involved actors are without any doubt the *users* of broadband Internet Access. They are producers of a huge amount of information used by third parties, and in some cases without awareness. Hence, Internet customers are not only clients but also products; their personal information is a good normally traded between companies. A proof is the widespread phenomenon of Web tracking [5]. As users employ an ISP to access services across the Internet, they expect the ISP to carry their data without looking inside the traffic and respecting their *privacy*. It is clear that there is a *trust* relationship between them, and an ISP is completely aware that lack of trust may bring customers to change provider, posing, in some cases, a serious threat of failure or bankruptcy for the ISP [6]. In case that a user does not trust her provider, she will likely change to another ISP, as witnessed by the great effort providers put in clients' loyalty programs.

On the other hand, Internet Access is just a means for users to enjoy services on the Internet. The entities that offer services on the Internet are called *content providers*: social networks, e-commerce portals, and search-engines are notable examples.

Few of them hold the majority of clients since the modern web is nowadays an oligopoly of few *big players* [7]. For instance, social networks converged around two main platforms, namely *Facebook* and *Twitter*, while search engines are dominated by *Google*. E-commerce sees a handful of major players like *Amazon*, *Ebay*, and *Alibaba*.

Although the relationship between users and content providers is clear and evident to the majority, ISPs and content providers are strictly connected by often conflicting regulations and directives, e.g., Net neutrality [8] force all stakeholders (e.g., ISP, content provider, governments, etc.) to treat all data on the Internet in the same way and without any discrimination for all users and services. However, a crucial voice of cost for an ISP is represented by the traffic outgoing from its network and, thus, it wishes to cache as much content as possible within its infrastructure (e.g., Comcast [9] throttled the data traffic uploaded by peer-to-peer applications without notifying the users). On the other hand, content providers want to have full and exclusive visibility on the behavior of their users, e.g., increasing the use of HTTPs (encrypted) in place of HTTP. Notice that the content of encrypted services cannot be cached using normal HTTP caches or proxies, leaving ISPs without a valuable means to reduce traffic to the exterior. For example, *Facebook* since April 2013 serves its content via encrypted connections: whereas it is a significant improvement from the privacy point of view, ISPs were certainly not happy for such decision. Also *Youtube* in January 2014 started to serve its content by means of encrypted connections, increasing significantly the amount of traffic not cacheable within ISP premises (*Youtube* traffic is more than 25% of total, see Fig. 4).

To partly reduce outgoing traffic, the trend shows ISPs hosting CDN nodes within their infrastructure. Whereas an ISP can benefit from hosting a CDN servers, few disadvantages are noticeable: the cache server must be powered, and especially must be filled with content coming from outside. The ISP has no control on the amount of content retrieved and whether only its customers are benefiting from that cache. Moreover, CDNs often use *encryption as default*, giving the ISP no visibility on users' behavior. Nevertheless, many works propose collaboration between ISPs and CDNs [10], [11]. Therefore, modern fashion in content delivering is certainly less appreciated by ISPs, since it decreases the amount of visible information at network level, but it is the only choice to deal with the big players of the Internet.

The last party involved in this ecosystem are those who can benefit from passive network measurements. Several actors are interested on information coming from the network; for an ISP it is important to know the clients' behavior to properly engineer its infrastructure (e.g., knowing which services are important for users allows to

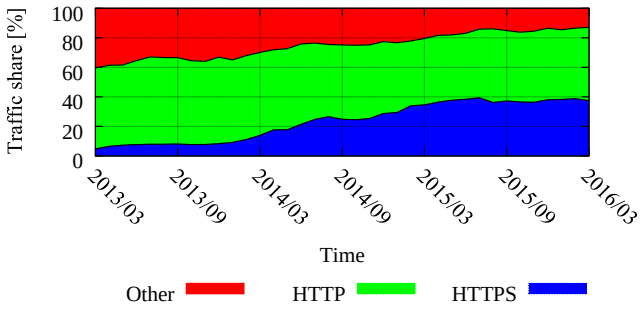

Fig. 2: Share of HTTP and HTTPs over last 3 years as measured on a PoP aggregating about 10,000 households from an European ISP

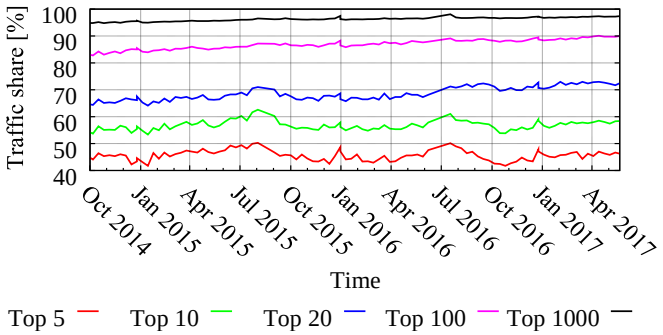

Fig. 3: Share of traffic due to the top-N domain names over the last three years. Top-{10,20,100} share exhibit an increasing trend.

better configure the network to carry such traffic). Moreover, information about customers' traffic unveils their interests, behavior and personal attitudes. The phenomenon of Web tracking is the proof that the brokerage of users' navigation data is widespread. Beside Web trackers, ISPs gather huge and complete customers' navigation data, typical regulatory frameworks to protect privacy are often indulgent and not up to date.[1]

In the next sections we focus our attention on such questions, keeping in mind that today's web is substantially different from 10 years ago. Collecting network measurement is getting harder due to encryption, Cloud/CDN infrastructure, and user's privacy concern. Nevertheless, what passing on the network is still a rich source of information for ISPs, researchers, and marketing enterprises. Many companies have as core business the collection of personal information to sell high detailed customers profile to other companies. This kind of business concerns marketing, advertising, political, and economy in general, the monetary turnover of such marketplace is exponentially increasing in recent years as well as users' awareness of privacy related issues [12].

III. Traffic Trend over the last years

Internet traffic in 2017 is very different from a decade ago. Many phenomena changed the ecosystem of the Web, causing substantial changes in traffic. The need for privacy and security boosted the adoption of encryption, with Transport Layer Security (TLS) being the king among secure protocols. Moreover, the Web converged around few big players, content providers that possess the majority of Web services, and can, alone, determine crucial and sudden changes on the global Internet traffic.

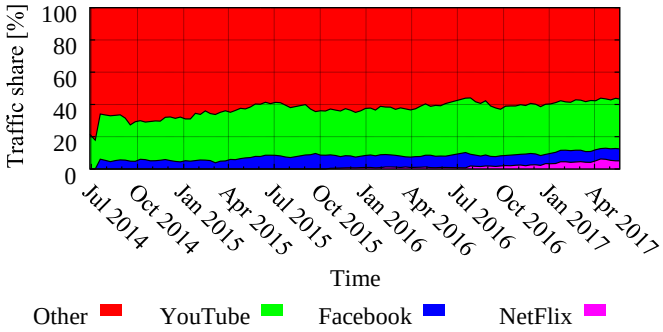

Fig. 4: Share of traffic due to three major Content Providers over the last three years

Fig. 2 shows the breakdown of the traffic generated by more than 10,000 Asymmetric Digital Subscriber Line (ADSL) customers over the last 3 years; a passive probe monitored all the traffic generated in a PoP of a European nationwide ISP, and created flow records using the Tstat passive monitoring tool [13]. It provides per-protocol breakdown and, in particular, the traffic share for HTTP, HTTPs, and other protocols. The portion of encrypted traffic raises from 10% in 2013 to 40 % in 2016, unveiling a 4 times increment just in 3 years. It is a proof of how the content providers and users care about privacy and security. If the share for HTTPs increased, HTTP did not considerably decrease. The room for HTTPs was made by all other protocols (red area in Fig. 2), showing a convergence of the Internet traffic on the two aforementioned protocols.

The Internet was born with a decentralized architecture, being a set of interconnected Autonomous Systems. Its original aim was to provide a simple means to deploy services and disseminate information, with the purpose of increasing the number of individuals able to provide their own contents to the world. However, in last years, the

[1] https://arstechnica.com/tech-policy/2017/04/trumps-signature-makes-itofficial-isp-privacy-rules-are-dead/

Internet is witnessing an unprecedented concentration around a handful of big players. To quantify this effect, we take advantage of Fig. 3, where we report the amount of traffic due to the top-*N* Fully Qualified Domains Names (FQDNSs) over the last three years. Each line represents the fraction of traffic due to the top-*N* FQDNs in terms of volume over the period of our dataset. We notice that the traffic due to the top-{10,20,100} FQDNs increased by more of 5% from 2014 to 2017, unveiling a concentration trend of the traffic around an always smaller set of Web services.

To further quantify this trend, we now focus on those content providers that offer the most pervasive Web Services in terms of users and traffic. To quantify the ascent of such giants, we analyze the volume of traffic generated by the most prominent ones, and report the results of the aforementioned analysis in Fig. 4. Profiting from the dataset introduced in the previous paragraph, we tracked the volume of traffic to *Facebook*, *YouTube*, and *NetFlix* over the last 3 years. We chose those three services as they are nowadays the leaders in their respective fields, and in terms of generated traffic. As depicted by the figure, their volume share constantly raised, unveiling the polarization of the Internet traffic around few online platforms. In particular, *YouTube* traffic raised from 25% (2014) up to more than 30% (2017), and *NetFlix*, available in the monitored country since 2015, accounts for more than 5% of volume.

IV. How Internet protocols break (or preserve) user's privacy

The aim of this section is to analyze how current Internet protocols protect user's privacy, and illustrate which information can an eavesdropper extract from passive monitoring of traffic. We separately analyze the network protocol layers, listing the most used protocols along with a discussion about which kind of knowledge can be extracted from eavesdropping their headers. In this work, we deal with the Internet traffic, so, we skip the discussion about data link layer (L2/OSI) issues. Indeed, L2 headers (Ethernet typically) are not propagated across routers but their lifespan is limited to a LAN.

A. Network layer

The Internet Protocol (IP) is the basis of the Internet. Among the fields contained in the header, Client and Server IP addresses are worthy of note. Client IP address is certainly the field carrying more sensible information; in the ISP network scenario, it uniquely identifies a particular user's home gateway. Even if home station addresses are often dynamic (i.e., managed and possibly changed by the network administrator), the association between them and customers is certainly retained by the ISP.

The Server IP address contains the remote endpoint contacted, assuming a scenario where the the ADSL customers are enjoying Web services located remotely on the Internet. The ISP can leverage this field to extract knowledge about which websites and services are used by a customer, and study her behavior. To hide this information, users can exploit anonymity network tools, with *TOR* being the king.[2] Nevertheless, with the advent of CDNs and Cloud Providers, the information obtainable by analyzing the server IP addresses is getting coarser. Indeed, many services are often co-located on the same servers owned by third party IaaS companies, with Akamai and Amazon being the leaders.

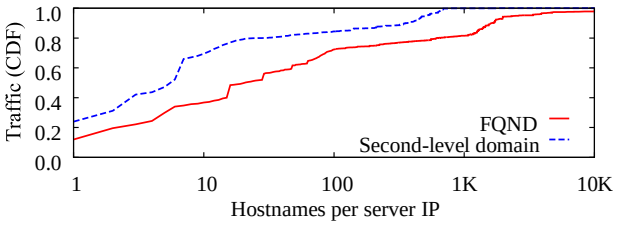

Fig. 5: CDF of the traffic related to server IP addresses with different numbers of domain names. One day of ISP traffic

Profiting from the dataset used in the previous section, we analyze one day of traffic in July 2017, and enumerate the domain names associated to each server. We notice that 33% of server IP addresses host more than one service – i.e., are contacted by monitored clients with more than one name. Servers having more than one hostname are responsible for a big fraction of the traffic. We quantify this in Fig. 5. For each server IP address, we count how many hostnames it holds and the number of bytes it handles. We then compute the fraction of traffic handled by servers having different number of names. Fig. 5 shows the resulting CDF. Less than 15% of the traffic is directed to servers having one hostname only. The picture only slightly changes when aggregating hostnames to the second-level – e.g., *apis.google.com* to *google.com*: only 33% of traffic is due to servers holding one second-level domain. That is, the server IP address lets an eavesdropper correctly infer the contacted domain name for less than 15% of the traffic; this increases up to 33% when taking into account only second-level domains. Part of the remaining traffic is necessarily misclassified, since many domain names (and thus services) run over the same servers.

Finally, a popular technique used to preserve privacy in IP-based network data is IP address anonymization. Several algorithms can be exploited to this end, with the most popular being Crypto-PAn [14], a cryptography-based sanitization tool. However, IP address anonymization still poses several challenges and questions [15], [16].

[2] https://www.torproject.org/

B. Transport layer

Transport layer builds on two widespread protocols, namely TCP and UDP. Their headers contain few fields, and poor personal information can be extracted by eavesdropping them. In particular, the only interesting fields are the port number, as they let a passive monitor to rebuild flows, and account each packet to the TCP/UDP connection it belongs to. Moreover, the server port number can give a hint about the network service contacted, as the list of well-known ports (0-1023) include popular services such as HTTP, TLS, or DNS.

C. Application Layer

Dozens of different application layer protocols there exists. However, nowadays the Web converged over two main protocols, namely HTTP and HTTPS; beside these, DNS still has a central role, being necessary to for reaching almost any Web server.

HTTP is the king of Web protocols, and it is used almost by all services on the Internet. It does not include encryption and, thus, all its headers are transmitted in clear. Therefore, all the details of HTTP transactions are offered to passive monitoring: an eavesdropper can read the URL and the full content of each document; moreover, additional headers such as Content-Type and User-Agent can provide meaningful information about the users' setup. Parameters of the webpages are transmitted in clear as well, and, thus, username and password might be extracted from packets.

To overcome privacy risks of HTTP, many services rely nowadays on HTTPS, that secures the former by putting it on the top of TLS. No HTTP header is transmitted in clear, leaving a passive monitor without any information about the underlying transaction. Nevertheless, TLS includes a field called Server Name Indication (SNI), where the client indicates in clear the domain name of the server being contacted.[3] Thus, the hostname of the contacted server is exposed to passive monitoring, unveiling in most cases the service accessed by the user [17]. Guidelines of TLS version 1.3 plan to encrypt this field, but this poses several technical challenges, as the handshake procedure would result more complicated and require additional RTTs before establishing the connection.[4]Moreover, many works showed the power of machine learning for extracting knowledge from encrypted connections, such as webpage URL [18] and users' QoE [19], [20].

The same considerations hold for the DNS protocol, where domain names are exchanged in plain text. Even DNSSEC does not guarantee confidentiality, but only provides origin authentication of the DNS data. DNS traffic can be used to provide fine-grained visibility even with encrypted traffic by rebuilding clients DNS cache and inspecting consequent TCP/UDP flows [21], [22].

Finally, to overcome the stiffness of TCP and TLS, some Content Providers designed and implemented their own protocol suites. This is the case of QUIC, designed by Google and implemented in Google Chrome, Android smartphones as well as on Google Web servers. It relies on UDP and provides authentication and confidentiality for HTTP transactions. Nevertheless, the server hostname is transmitted in clear, posing the same privacy issues of TLS. Moreover, QUIC transmits client's User-Agent in clear, unveiling user's device type.[5] In addition to Google, Facebook designed its own application layer protocol, called Zero.[6] It is based on QUIC's crypto module, but relies on TCP instead of UDP. It is implemented in Facebook and Instagram Web servers as well as on the mobile applications of such platforms. Being by design similar to QUIC, the same considerations hold: the server hostname is exchanged in clear, leaking to passive monitors the name of the contacted service.

V. Ethical issues

Given what has been discussed so far in this work, we can identify a set of privacy and ethical issues. In the following, we will discuss these issues from two main points of view: the user one and the ISP one.

From the ISP point of view, we can summarize the issue with the question: "Is it right to access the data exchanged by a user?". It is not straightforward to give an answer to this question: in fact, one could instinctively be tempted to respond:
"no, it is wrong at all" because of the user's privacy, but there are some other aspects that must be considered. In fact, it seems obvious that when a user relies on an ISP to access the Internet. She should also trust the ISP and assume that her data pass through ISP network without any inspection. On the other side, if an ISP could inspect user traffic, at least to apply some Quality of Service (QoS) policy or just some internal routing optimizations, it would be able to improve the user network experience. Given that, the ISP could be justified to monitor the users' traffic but where is the boundary located and where the ISP should stop in the inspection. These questions come out from the fact that the current digital world most advantageous activity is represented by big data analysis and users profiling [23].

Therefore, if an entity can access a huge amount of information from a large number of users it might be tempted to sell these data to the highest bidder, that is exactly the case of an ISP. It seems useful to try to define a limit at which the inspection is deep enough for the ISP optimizations and not

[3] https://tools.ietf.org/html/rfc3546

[4] TLS 1.3 Draft Specifications are available at https://github.com/tlswg/tls13spec

[5] QUIC specifications available at https://goo.gl/SBS95v

[6] https://code.facebook.com/posts/608854979307125/building-zeroprotocol-for-fast-secure-mobile-connections/

too intrusive for users' privacy and security. It is not evident in which way the traffic payload inspection could be interesting for routing optimization purposes and then a solution could be to force ISPs to inspect only transport information (e.g., IP addresses, TCP/UDP ports). Nevertheless, TCP and DNS analysis are enough to assess websites a client is visiting. The payload inspection, on the other hand, could be useful to prevent malicious intents or illegal communications. As each ISP provides the network access point for its end-users, that are typically a significant amount, it could be the best point to detect criminal activity, identify involved parties, and prevent them. This fact seems to justify a deeper traffic inspection by the ISPs but, in order to protect users' privacy, the inspection activity could be delegated to a trusted third party such as police or governments institutions[7]. It is clear how the decision varies considering different aspects of the topic and how difficult could be to take a decision about it.

On the other hand, from the end user point of view the issue is mainly a matter of privacy and personal information disclosure. In this case, we identified one main question: "the user's privacy must be managed only by the user himself?" In other words, should the end-user care about his privacy while considering all the rest of the world as untrusted or there should be some privacy level guaranteed by the ISPs? In this case we believe that truth lies somewhere in the middle: ISPs should implement their services in order to not disclose users' information and end-users should take care of their personal data when sent on the network. Currently, personal data protection involves data encryption that is a crucial ethical and legal point of discussion. Recent events have underlined how encryption can cause problems in case of investigation against criminal acts, e.g., phones used by terrorists that are totally ciphered and that the authorities cannot access (that is what they publicly said) [24]. The ethical discussion about data ciphering can be very hard because it could be impossible to decide what is right among for main possible decisions.

However, the ethical discussion about encryption is out of the scope of this work. We expect a great debate about this topic in the near future, where the role of network will take second place, as major Content Providers will be involved; *Whatsapp* decision to encipher all messages and hiding them from inspection of anyone is a good example and will certainly have many consequences [25].

For the ISP case, data ciphering is a crucial point: as we described in Sec. IV data anonymization could limit the ISP information about the user traffic but the sensitive data can also be accessed from the payload (which often is not obfuscated at all). Then, it is clear that the user has to take care of his data by ciphering payload when it is needed for privacy means but it is not clear how to address the social security problem tied to ciphered data. Moreover, we claim that even when having encrypted connections (i.e. HTTPs) some information is still assessable (e.g., website history).

In conclusion, privacy and ethical issues of the ISPs in the modern web can be placed in the current digital world ethical discussion: it is about finding a trade-off between personal privacy a public security.

VI. Alternative scenario

As presented in the preceding sections, actually users have typically a high level of privacy on the payload level, but a lower one on the network level. In other words, ISPs can only know if, when, and how many times a user have visited a particular website, without having access of what activity the user have performed on it. How to conciliate the need of privacy for the user, with the technical (i.e., QoS) and commercial needs for the ISPs.

A possible solution might be an ecosystem where users can voluntarily permit their ISP to access the payload of their communications. In exchange the users can enjoy globally a better service, thanks to QoS: typically, we have different expectations on the waiting time to access a resource, based on the type of content we are accessing; for example, waiting some (not too much) time to access a textual web page is not a big deal, whilst having a video stream that stops constantly for buffering can be really annoying. Using QoS, ISPs can fine tune the bandwidth given to the users, in function of the accessed resources.

Users may also have economic benefits: they can share the revenues that the ISP obtained by selling their data, in terms of a reduction of their monthly fee for Internet connection. This can be also a big boost in giving access to Internet in the growing countries, and for the most disadvantaged people in first world countries: lowering the fee for accessing Internet, more people can gain access to it, thus giving more data to the ISP for them to sell, creating a virtuous cycle.

However, the depicted alternative poses issues from an ethical point of view. In particular, it must be ensured that users must not be forced to "sell" their data to the ISPs. The risk is that ISP increase the fee to the users that do not want to give their data to them, thus making the choice practically obligatory for almost everyone. This could be avoided only by a strict control of the fees by the government regulatory agencies, thus this could be a problem in some countries, where this could be seen as an unacceptable intervention of the government in the private economy sector or even a means for political control. Legislating about Web users' privacy is hard, and might lead to resounding failures, as witnessed by the many concerns about 209 European ePrivacy Directive [26], [27], [28]. The ISPs can also force users to give access to their data by using the QoS, slowing Internet access to users not willing to "sell" their data. This behavior by ISP, although avoidable from the legal point of

[7] In that case, it should also be discussed the trustworthiness of that entities and identified a boundary between public security and a big brother effect.

view (i.e. Service Level Agreement), can be hardly traceable, thus creating an "informal" threat to the users' liberty.

Concluding, we can say that the alternative scenario proposed is feasible, but it should be deployed with great attention for the users' rights, and with great scrutiny regarding the ISPs behavior.

VII. Conclusions

Nowadays, ethical and privacy issues in the networking field are getting more important. Internet measurement fields raise ethical inquiries when it comes to privacy and security concerns of individual users using the Internet. In particular, there is a consistent debate about how ISPs should collect and treat network measurements. This kind of information are fruitful source of knowledge from multiple points of view: research, traffic engineering, governmental, and e-commerce can benefit from measurements retrievable through inspection of network traffic. The data collected in modern web might carry personal information about the users exposed to monitoring, and generates critical ethical and moral issues.

The aim of this work is to shed light on the ethical aspect of the information exposed to ISPs in the modern web. We highlight ethical issues deriving by the current situation and provide general guidelines and best practices to cope with the collection of network traffic measurements.

References

[1] Kjell Brunnstrom, et al. Qualinet White Paper on Definitions of Quality of Experience, March 2013, Novi Sad.

[2] B. Zevenbergen, I. Brown, J. Wright, and D Erdos. Ethical privacy guidelines for mobile connectivity measurements. Oxford Internet Institute, University of Oxford, 2013.

[3] Luca Vassio, Hassan Metwalley, and Danilo Giordano. *The Exploitation of Web Navigation Data: Ethical Issues and Alternative Scenarios*, pages 119–129. Springer International Publishing, Cham, 2016.

[4] R. Hofstede, P. eleda, B. Trammell, I. Drago, R. Sadre, A. Sperotto, and A. Pras. Flow monitoring explained: From packet capture to data analysis with netflow and ipfix. *IEEE Communications Surveys Tutorials*, 16(4):2037–2064, Fourthquarter 2014.

[5] H. Metwalley, S. Traverso, and M. Mellia. Using passive measurements to demystify online trackers. *Computer*, 49(3):50–55, Mar 2016.

[6] Jyh-Shen Chiou. The antecedents of consumers loyalty toward internet service providers. *Information & Management*, 41(6):685–695, 2004.

[7] Chris Anderson and Michael Wolff. The web is dead. long live the internet. 2010.

[8] Tim Wu. Network neutrality, broadband discrimination. *J. on Telecomm. & High Tech. L.*, 2:141, 2003.

[9] Peter Svensson. Comcast Blocks some Subscriber Internet Traffic. http: //www.nbcnews.com/id/21376597/#.WV9a-SdLdpg, 2009.

[10] Benjamin Frank, Ingmar Poese, Yin Lin, Georgios Smaragdakis, Anja Feldmann, Bruce Maggs, Jannis Rake, Steve Uhlig, and Rick Weber. Pushing cdn-isp collaboration to the limit. *ACM SIGCOMM Computer Communication Review*, 43(3):34–44, 2013.

[11] Dongmyung Lee, Jeonghoon Mo, and Jinwoo Park. Isp vs. isp+ cdn: can isps in duopoly profit by introducing cdn services? *ACM SIGMETRICS Performance Evaluation Review*, 40(2):46–48, 2012.

[12] Sue Conger, Joanne H Pratt, and Karen D Loch. Personal information privacy and emerging technologies. *Information Systems Journal*, 23(5):401–417, 2013.

[13] A. Finamore, M. Mellia, M. Meo, M. M. Munafo, P. D. Torino, and D. Rossi. Experiences of internet traffic monitoring with tstat. *IEEE Network*, 25(3):8–14, May 2011.

[14] Jun Xu, Jinliang Fan, M. H. Ammar, and S. B. Moon. Prefix-preserving ip address anonymization: measurement-based security evaluation and a new cryptography-based scheme. In *Network Protocols, 2002. Proceedings. 10th IEEE International Conference on*, pages 280–289, Nov 2002.

[15] Ron A. Dolin. Search query privacy: The problem of anonymization. pages 280–289, April 2010.

[16] Stevens Le Blond, David Choffnes, Wenxuan Zhou, Peter Druschel, Hitesh Ballani, and Paul Francis. Towards efficient traffic-analysis resistant anonymity networks. *SIGCOMM Comput. Commun. Rev.*, 43(4):303–314, August 2013.

[17] W. M. Shbair, T. Cholez, J. Franois, and I. Chrisment. Improving sni-based https security monitoring. In *2016 IEEE 36th International Conference on Distributed Computing Systems Workshops (ICDCSW)*, pages 72–77, June 2016.

[18] Roberto Gonzalez, Claudio Soriente, and Nikolaos Laoutaris. User profiling in the time of https. In *Proceedings of the 2016 Internet Measurement Conference*, IMC '16, pages 373–379, New York, NY, USA, 2016. ACM.

[19] Giorgos Dimopoulos, Ilias Leontiadis, Pere Barlet-Ros, and Konstantina Papagiannaki. Measuring video qoe from encrypted traffic. In *Proceedings of the 2016 Internet Measurement Conference*, IMC '16, pages 513–526, New York, NY, USA, 2016. ACM.

[20] Martino Trevisan, Idilio Drago, and Marco Mellia. Pain: A passive web speed indicator for isps. 2017.

[21] Ignacio N Bermudez, Marco Mellia, Maurizio M Munafo, Ram Keralapura, and Antonio Nucci. Dns to the rescue: discerning content and services in a tangled web. In *Proceedings of the 2012 ACM conference on Internet measurement conference*, pages 413–426. ACM, 2012.

[22] Tatsuya Mori, Takeru Inoue, Akihiro Shimoda, Kazumichi Sato, Keisuke Ishibashi, and Shigeki Goto. Sfmap: Inferring services over encrypted web flows using dynamical domain name graphs. In *International Workshop on Traffic Monitoring and Analysis*, pages 126–139. Springer, 2015.

[23] Hsinchun Chen, Roger HL Chiang, and Veda C Storey. Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 36(4):1165–1188, 2012.

[24] Us police groups claim encryption has made iphones the 'device of choice' for criminals. *Reuters and VICE News*, 2016.

[25] Amy Nordrum. In privacy versus security, end-to-end encryption is definitely winning. *IEEE Spectrum*, 2016.

[26] Bert-Jaap Koops. The trouble with european data protection law.
International Data Privacy Law, 4(4):250–261, 2014.

[27] Christina Markou. Behavioural advertising and the new eu cookie lawas a victim of business resistance and a lack of official determination. In *Data Protection on the Move*, pages 213–247. Springer, 2016.

[28] Ronald Leenes and Eleni Kosta. Taming the cookie monster with dutch law–a tale of regulatory failure. *Computer Law & Security Review*, 31(3):317–335, 2015.